

Combining vertex and soft electron b-tagging methods using a neural network approach

Marcin Wolter

Tufts University

On leave from the H. Niewodniczanski Institute of Nuclear Physics,
Krakow, Poland

1 December 2000

On my WWW page: www.tufts.edu/~mwolte01





Combining b -tagging methods using a neural network approach ¹

Marcin Wolter and Anna Kaczmarska
Institute of Nuclear Physics
30-059 Kraków, ul. Kawczy 26a, Poland

Abstract

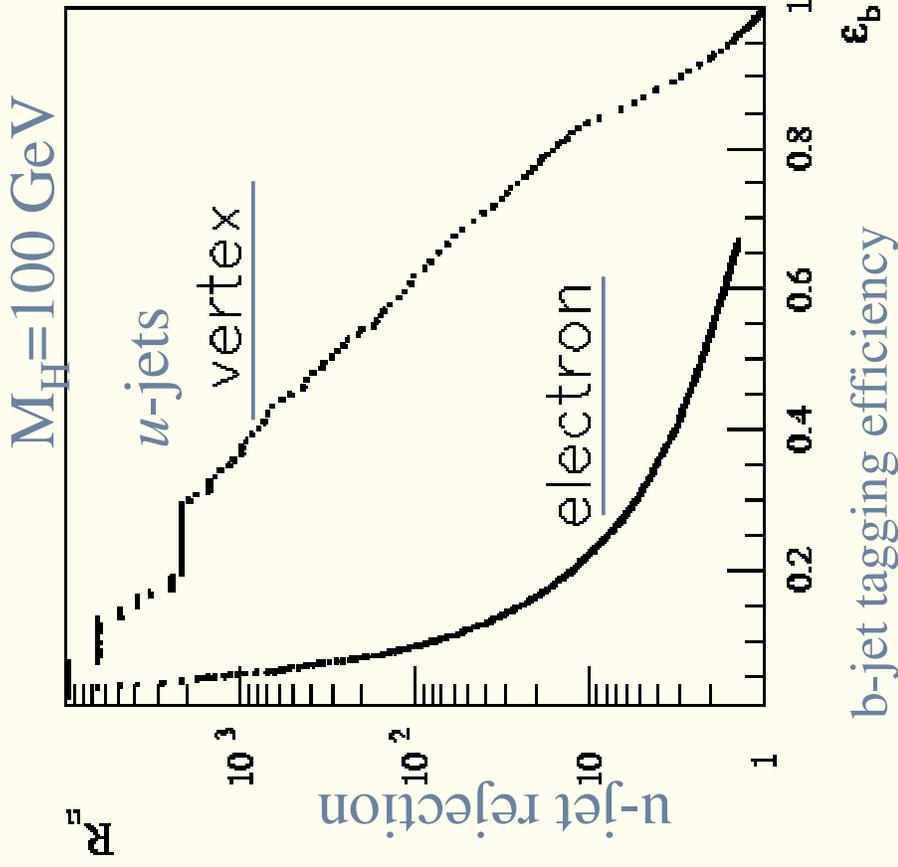
The most powerful method of b -jet tagging in ATLAS experiment is based on the relatively long lifetime of B -hadrons and therefore a separation of the B -hadron decay vertex from the pp interaction point [1]. The efficiency of the method is improved while combined with another tagging algorithm using low- p_T electrons from semileptonic b -quark decays [2]. For the first time the neural network approach is used to combine both tagging algorithms. Rejection of background u -jets is improved by 16% and c -jets by 10% comparing to the pure vertex method.

M. Wolter and A. Kaczmarska

ATL-INDET-2000-023

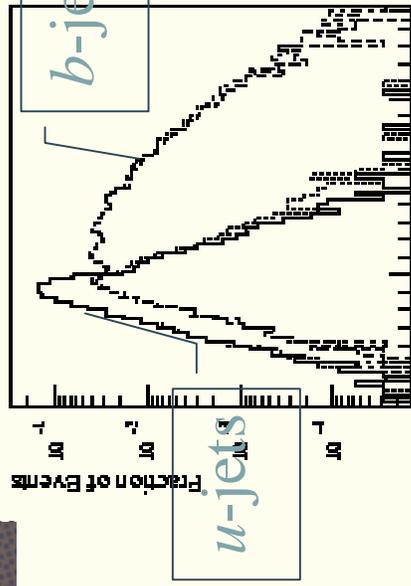
¹Work supported in part by Polish Government grants 2 P03B 121 18, 2 P03B 118 19 and 620/E-77/SPUB/CERN/P-03/DZ 2/99.

Two b -tagging methods at ATLAS

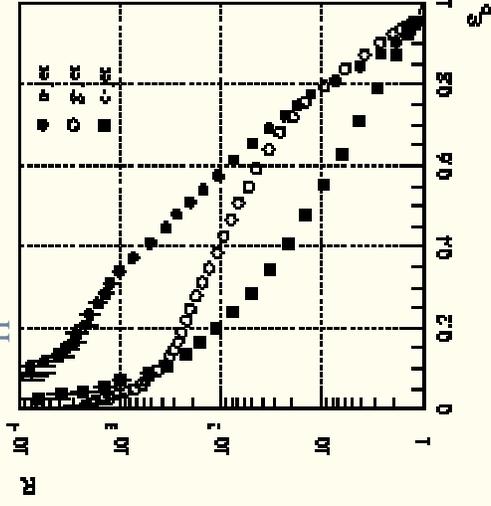


- # Combining two b -jet tagging methods based on soft electron and long decay path of b mesons.
- # $\text{BR}(b \rightarrow e) \approx 17\%$ only!
- # Combination of “good” and “poor” methods.
- # Background: u -, c - and g -jets.

Vertex tagging



$M_H = 400 \text{ GeV}$



Based on long B-mesons lifetime ($c\tau \approx 470 \mu\text{m}$).

Weight for single track based on the signed impact parameter (the track crosses the jet axis in front or behind the primary vertex).

Reconstruction on R - ϕ plane.

Significance: $S_i = d_0 / \Delta(d_0)$

Weight: $r_i = f_b(S_i) / f_u(S_i)$

Jet weight – the sum of the logarithms of the single track weights.

Performance of the vertex *b*-tagging method

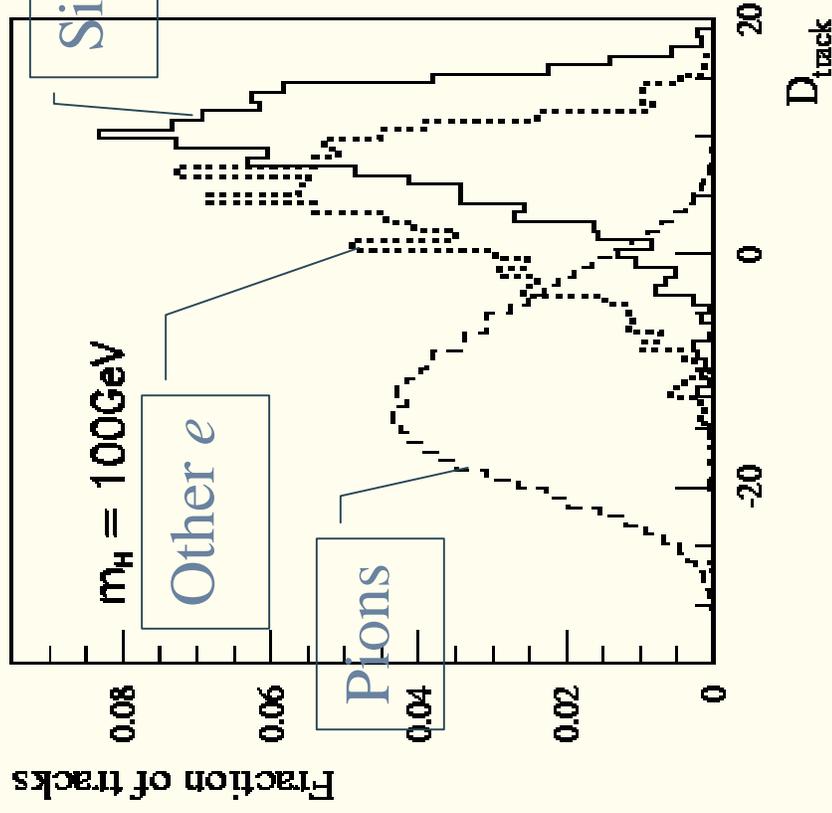
$$\varepsilon_b = 50\%$$

$$m_H = 100 \text{ GeV} \qquad m_H = 400 \text{ GeV}$$

	xKalman	iPatRec	xKalman	iPatRec
R_H	326 ± 37	391 ± 49	126 ± 9	183 ± 17
R_g	135 ± 12	148 ± 14	59 ± 3	58 ± 3
R_c	13.6 ± 0.4	11.7 ± 0.3	13.3 ± 0.4	13.4 ± 0.4

Published in *ATLAS TDR CERN/LHCC/99-14*

Soft electron tagging



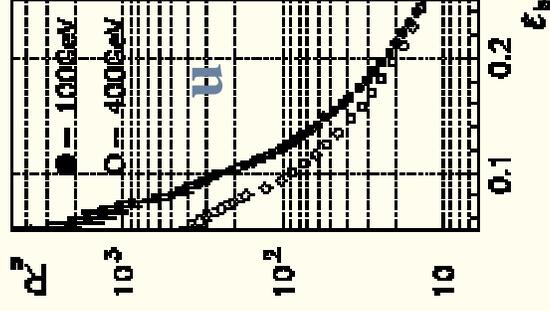
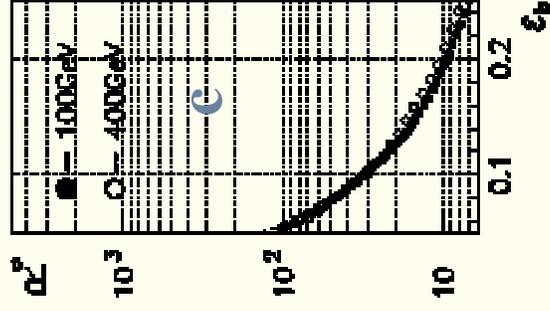
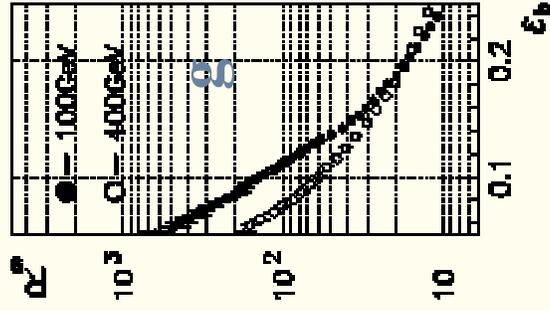
- # Based on soft electron identification.

- # Nine variables (inner detector + calorimeter) used to identify electron track.

- # Jet weight defined by the weight of the most “electron-like” track.

- # Limited by $\text{BR}(b \rightarrow e) \approx 17\%$.

Soft electron tagging performance



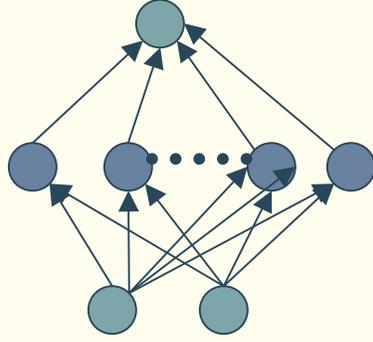
m_H [GeV]	100	400
\mathcal{E}_b^{soft}	8.2%	12.3%
R_g	209 ± 34	43 ± 3
R_u	470 ± 92	55 ± 3
R_c	44 ± 3	22 ± 1

ATLAS-PHYS-2000-012

A.Kaczmarska, M.Wolter ‘Improving overall b -jet identification with soft electron tag and neural network approach’, submitted to Acta Physica Polonica

Designing a neural network

Two inputs, one output.

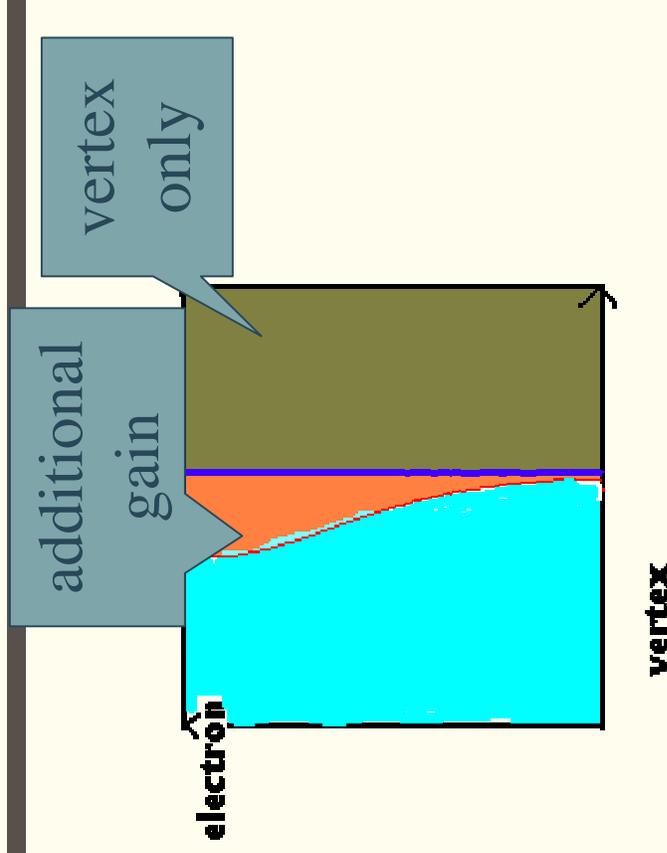


One hidden layer.

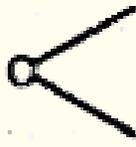
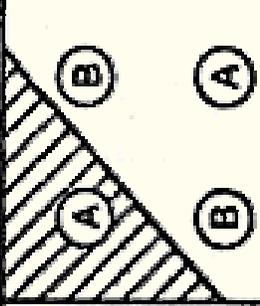
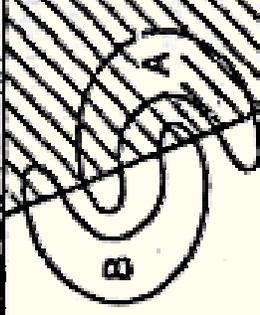
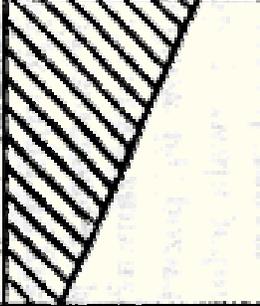
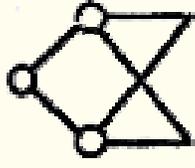
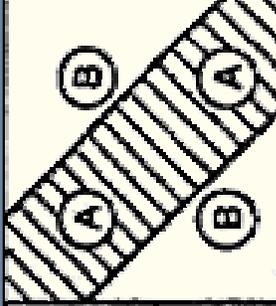
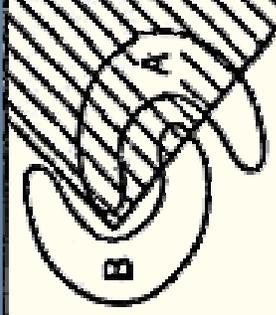
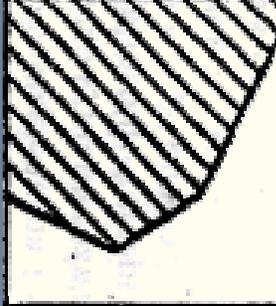
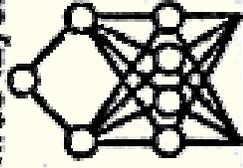
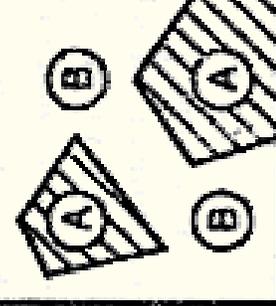
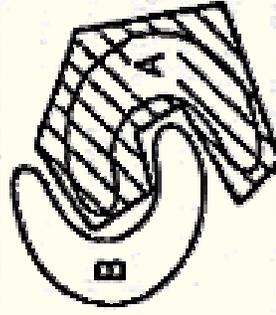
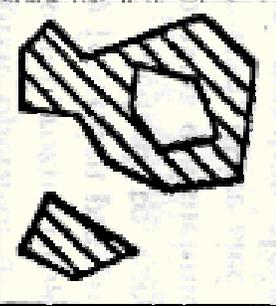
Smooth output function – activation logistic.

Simple feedforward network

Stuttgart Neural Network Simulator used.



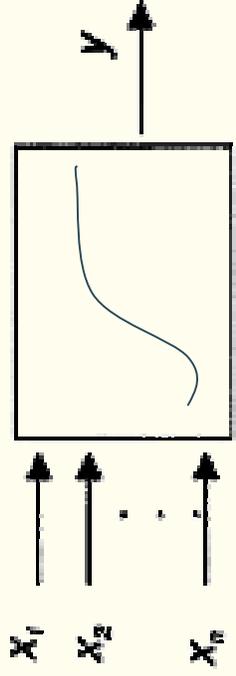
Why one hidden layer?

Structure	Type of Decision Regions	Exclusive-OR Problem	Classes with Mesned Regions	Most General Region Shapes
Single-layer 	Half plane bounded by hyperplane			
Two-layers 	Convex open or closed regions			
Three-layers 	Arbitrary (Complexity limited by number of nodes)			

one hidden layer

two hidden layers

Logistic activation function



A neuron - many inputs, one output

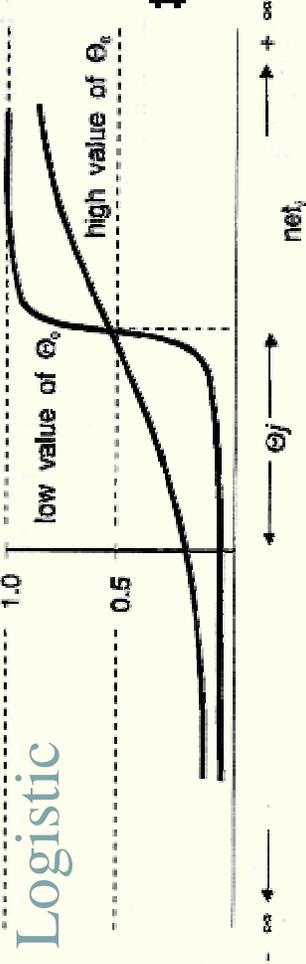
$y = f(w_1 \cdot x_1, \dots, w_n \cdot x_n)$

Where f is an arbitrary function: linear, step or

logistic:
$$y = \frac{1}{1 + \exp(-\theta_0 \cdot net_j)}$$

$$net_j = \theta_j + \sum_{i=1}^n w_i \cdot x_i$$

w_i and θ_j are tuned during the learning process.



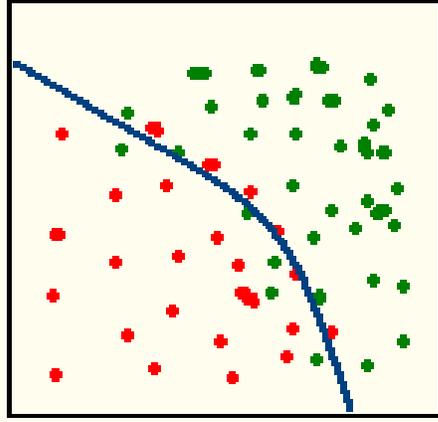
Why Stuttgart Neural Network Simulator?

- # Free university software available for many platforms.
- # Online monitoring of learning process (χ^2 based)
- # Plots of output (any node) vs inputs of two chosen nodes.
- # Easy to built and train networks – graphics interface.
- # Good documentation.

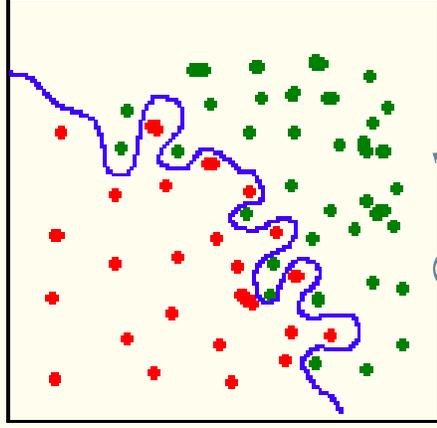
Learning procedure

- # Training sample: b -jets (signal) 9618 events, u -, c -, g -jets (background) $3 \cdot 4000$ events.
- # Verification sample: same proportions of signal and background, but independent data sets.
- # χ^2 for both samples verified during learning process to avoid *overlearning*.

Overlearning

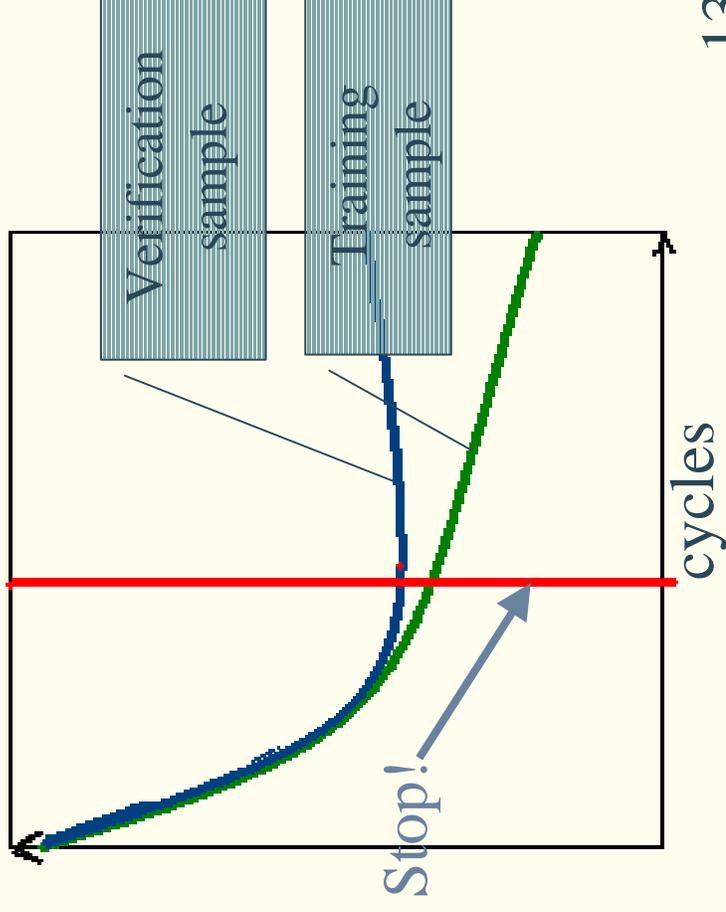


Proper

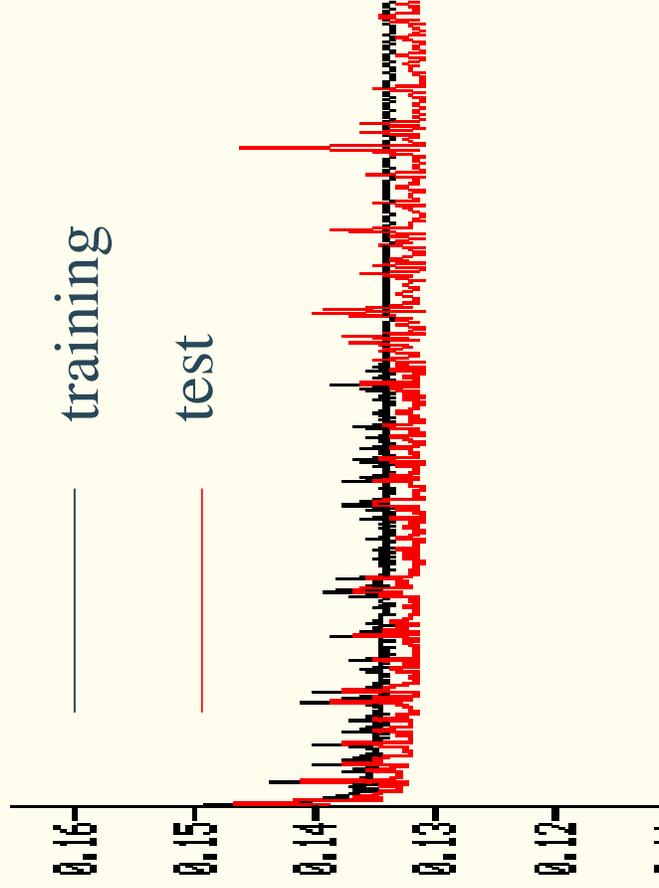


Overlearned

Overlearning - network learns single events instead of general characteristics.

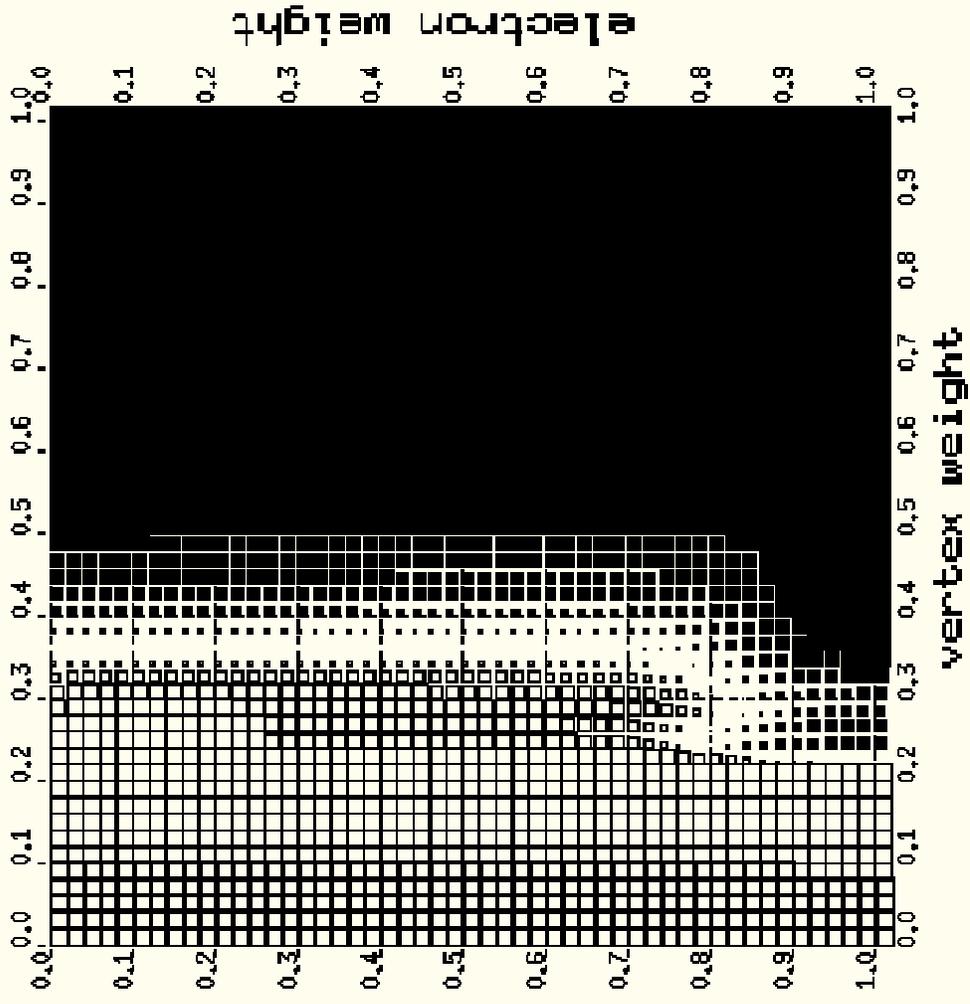


Learning...



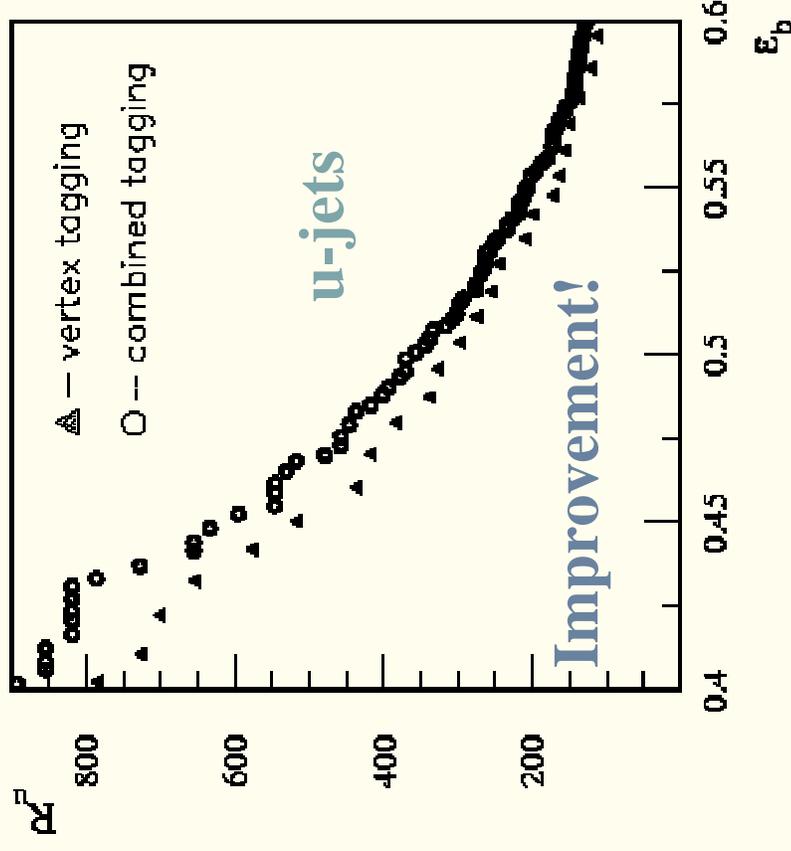
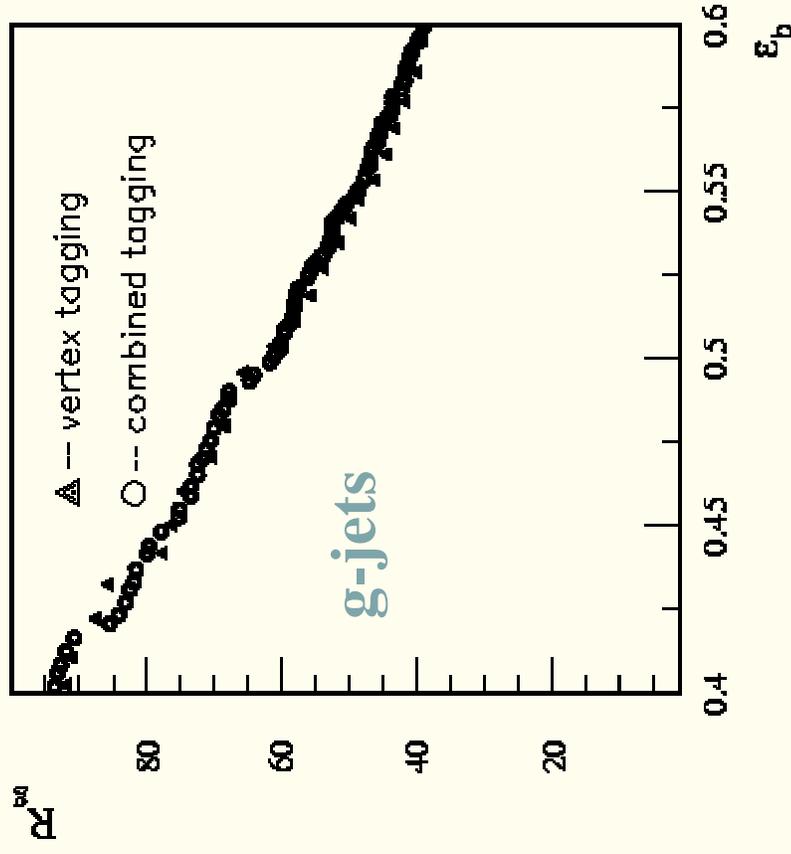
- # No sign of overlearning.
- # Difference due to slightly different statistics.
- # 24 nodes in the hidden layer – best performance, still no overlearning (but similar performance for different number of nodes in the range 8-48).

Weight function

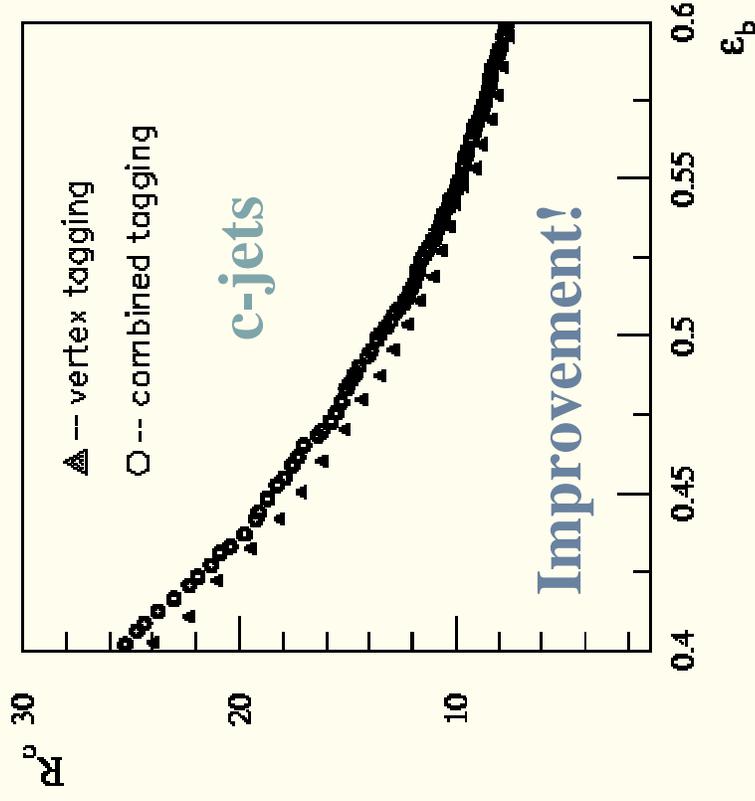


Function shape agrees with our intuition.

Improvement of background rejection

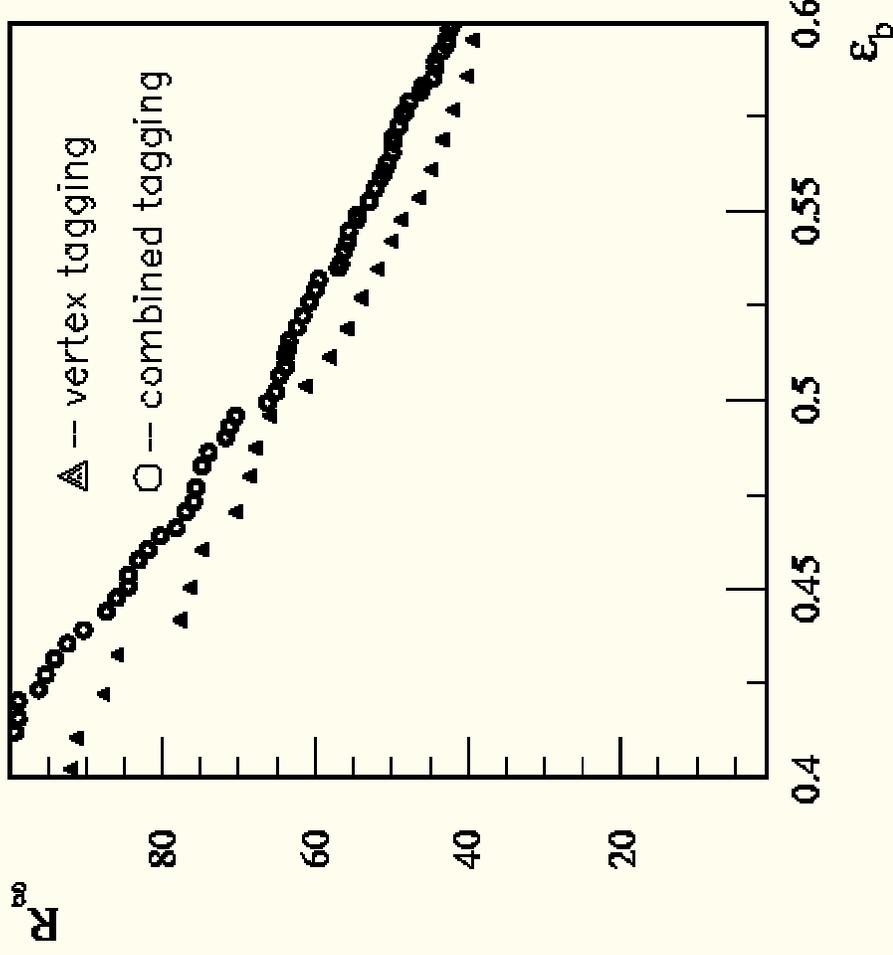


Improvement of background rejection



- # Improved rejection against μ - and c -jets.
- # NN allows to combine two tagging methods.

Network trained to reject g-jets



Network can be specially trained against g-jets (but worse rejection against another background sources).

Results

	vertexing rejection (at 50% eff.)	combined rejection (at 50% eff.)	rejection improvement
<i>g</i> -jets	61	61	none
<i>u</i> -jets	298	345	16%
<i>c</i> -jets	12.3	13.5	10%

	rejection improvement
<i>g</i> -jets	10%
<i>u</i> -jets	1%
<i>c</i> -jets	1%

Combinatorial method:

ATLAS TDR CERN/LHCC/99-14

Different vertexing rejections comparing to ATLAS TDR due to different data samples and track preselection criteria.

CDF experiment

- # ATLAS future silicon tracker has better performance than CDF vertex detector.
- # Combining vertex and soft lepton b -tagging methods in CDF should improve b -tagging performance more significantly than in ATLAS.
- # Tufts group is planning to perform similar analysis using CDF Monte Carlo and data.

Conclusions

- # Neural network algorithm allows efficient combination of two very different tagging methods.
- # CDF should benefit more than ATLAS from combining vertex and soft lepton tagging algorithms.